

Extra Chance Generalized Hybrid Monte Carlo[☆]

Cédric M. Campos*

*Dept. Matemática Aplicada e IMUVA, Fac. Ciencias, UVA
Paseo de Belén 7, 47011 Valladolid, Spain*

J. M. Sanz-Serna

*Dept. Matemáticas, Universidad Carlos III de Madrid
Edificio Sabatini, 28911 Leganés (Madrid), Spain*

Abstract

We study a method, Extra Chance Generalized Hybrid Monte Carlo, to avoid rejections in the Hybrid Monte Carlo method and related algorithms. In the spirit of delayed rejection, whenever a rejection would occur, extra work is done to find a fresh proposal that, hopefully, may be accepted. We present experiments that clearly indicate that the additional work per sample carried out in the extra chance approach clearly pays in terms of the quality of the samples generated.

Keywords: sampling methods, hybrid Monte Carlo, detailed balance, delayed rejection, rejection avoidance, molecular dynamics

PACS: 02.50.Ng, 02.70.Ns, 02.70.Tt, 05.10.Ln, 31.15.xv

2010 MSC: 60J22, 65C05, 65C40, 74A25

1. Introduction

In this paper we study a technique, Extra Chance Generalized Hybrid Monte Carlo (XCGHMC), to avoid rejections in the Hybrid Monte Carlo (HMC) method [8] and its variants (e.g. [3, 9, 15–17]). In the spirit of what in the statistics literature is called *delayed rejection* [12, 20, 32], whenever a rejection would occur, additional work is done to find a fresh proposal that, hopefully, may be accepted. Rejections, while essential to ensure that the algorithms sample from the right target probability distribution, contribute to an increase of the correlation of the samples [22, 31]. Furthermore, for algorithms with partial momentum refreshments [15, 17], each rejection necessarily requires a flip of the momentum and interferes with the underlying Hamiltonian dynamics. We present experiments that clearly indicate that the additional work per sample carried out in the extra chance approach in order to avoid rejections clearly pays in terms of the quality of the samples generated.

The algorithm studied here is mathematically equivalent to that suggested by Sohl-Dickstein, Mudigonda and DeWeese in [30]. However the actual formulas for the acceptance probability in the present work are

[☆]A MATLAB implementation of the algorithm and other utilitarian scripts can be found at <http://github.com/vitaminace33/xhmc>.

*Corresponding author

Email addresses: cedricmc@{uva, icmat}.es (Cédric M. Campos), jmsanzserna@gmail.com (J. M. Sanz-Serna)

different from those in [30]. We believe that the formulas used here provide insight into the probabilities involved. Furthermore we prove that the extra chance algorithm actually satisfies detailed balance/stochastic reversibility; this is surprising, as [30] suggests that the avoidance of rejections comes at the price of the violation of detailed balance. Detailed balance is of course a valuable property in the performance and analysis of Markov Chain Monte Carlo algorithms [24]. For instance the estimation of the effective sample size or autocorrelation time used in our experiments (taken from [10]) relies on the chain being reversible with respect to the target distribution.

In turn the basic idea in [30] is not essentially different from delayed rejection [12, 20, 32]. However some peculiarities of HMC (momentum flip, modified detailed balance vs. standard detailed balance, proposals being obtained via a deterministic flow) make it difficult, or even impossible, to apply the material of [12, 20, 32] to devise or analyze algorithms similar to those considered in [30] or here.

The literature is not lacking in suggestions to avoid rejection/momentum flips in HMC, see e.g. [1–3, 29, 33]. Comparisons between those techniques and the extra chance approach are not within our scope here. Neither shall we be concerned with comparing HMC with alternative sampling algorithms as done in [7].

In Section 2 we present the extra chance algorithm. Section 3 is devoted to an analysis of the acceptance probabilities. Proof of concept numerical experiments are reported in Section 4 and Section 5 concludes. Some more mathematical results are given in Appendices A–C.

2. Algorithm

The aim is to obtain samples x_n from a target probability distribution in the *state space* \mathbb{R}^d with density

$$\frac{1}{Z} \exp(-\beta V(x)), \quad Z = \int_{\mathbb{R}^d} \exp(-\beta V(x)) dx. \quad (1)$$

The algorithms considered here do not require that Z be known; they merely need to evaluate V and its gradient ∇V . They generate a Markov chain $x_0 \rightarrow x_1 \rightarrow \dots \rightarrow x_N$ [24] that has (1) as an invariant distribution, in such a way that under suitable ergodic behavior, it is possible to estimate averages $\langle A \rangle$ with respect to the target by taking means of the values of A along a realization of the chain:

$$\langle A \rangle = \frac{1}{Z} \int_{\mathbb{R}^d} A(x) \exp(-\beta V(x)) dx \approx \frac{1}{N+1} \sum_{n=0}^N A(x_n). \quad (2)$$

The format in (1) implies that the density is everywhere positive. As shown in e.g. [9], it is not difficult to extend the algorithms and analysis to cases where the density vanishes in a subset of the state space.

Regardless of the details of the application in mind, HMC and its variants use the Hamiltonian formalism of classical mechanics. The components of x are interpreted as generalized co-ordinates describing the configuration of a mechanical system and an auxiliary d -dimensional vector y is introduced whose components represent the associated conjugated momenta. We shall use the notations $z = (x, y)$, $z' = (x', y')$, etc. to refer to points in the *phase space* $\mathbb{R}^d \times \mathbb{R}^d$. If M is a user-specified, symmetric positive-definite $d \times d$ mass matrix, the algorithms use the Hamiltonian function (total mechanical energy)

$$H(z) = \frac{1}{2} y^T M^{-1} y + V(x), \quad (3)$$

and the probability distribution in the phase space defined by the unnormalized density

$$\rho(z) = \exp(-\beta H(z)) = \exp(-(\beta/2) y^T M^{-1} y) \times \exp(-\beta V(x)). \quad (4)$$

```

00 given  $z = (x, y) \in \mathbb{R}^d \times \mathbb{R}^d$ 
01  $y \leftarrow y - (\Delta t/2)\nabla V(x)$ 
02 for  $\ell = 1$  to  $L - 1$ 
03      $x \leftarrow x + \Delta t M^{-1}y$ 
04      $y \leftarrow y - \Delta t \nabla V(x)$ 
05 end for
06  $x \leftarrow x + \Delta t M^{-1}y$ 
07  $y \leftarrow y - (\Delta t/2)\nabla V(x)$ 
08  $(x', y') = (x, y)$ 

```

Figure 1: Velocity Verlet integrator for advancing $L \geq 1$ time-steps of length $\Delta t > 0$ starting from the initial point z and producing the final point $z' = \mathbf{I}(z)$.

The product structure of ρ implies that x and y are stochastically independent; x is distributed according to the target (1) and $y \sim \mathcal{N}(0, M)$, i.e. y is Gaussian with zero mean and covariance matrix M .

The dynamics associated with (3) is given by

$$\frac{d}{dt}x = M^{-1}y, \quad \frac{d}{dt}y = -\nabla V(x), \quad (5)$$

a system of differential equations whose solution flow exactly preserves the distribution (4), see e.g. [26]. In practice this flow cannot be computed in closed form and one has to resort to numerical approximations; the Störmer-Verlet/leapfrog integrator [13, 18, 27, 28] is the method of choice. Figure 1 shows pseudocode for computing the result $z' = (x', y') = \mathbf{I}(z)$ of L integration time-steps of length $\Delta t > 0$ starting from the initial point z . The transformation \mathbf{I} , which maps the phase space into itself, is both *volume preserving* (i.e. has unit Jacobian determinant) and *reversible*. Reversibility means that, for each z' , $\mathbf{I}^{-1}(z') = \mathbf{F}(\mathbf{I}(\mathbf{F}(z')))$, where \mathbf{F} denotes the *momentum flip transformation*: $\mathbf{F}(x, y) = (x, -y)$. Thus to recover the initial point $z = \mathbf{I}^{-1}(z')$ of a numerical integration it is sufficient to flip the momentum, take L (forward) time-steps and flip again the momentum; this exactly matches a fundamental property of the true solution flow of (5). Note that \mathbf{F} is volume preserving and, furthermore, $H(z) = H(\mathbf{F}(z))$.

The XCGHMC algorithm is summarized in Fig. 2. The proof that ρ in (4) is an invariant density of the chain $z_0 \rightarrow z_1 \rightarrow \dots \rightarrow z_N$ is given in Appendix A. Appendix B shows that, if $y_0 \sim \mathcal{N}(0, M)$, the marginal chain $x_0 \rightarrow x_1 \rightarrow \dots \rightarrow x_N$ obtained by taking the x components satisfies detailed balance with respect to the target distribution (1). As pointed out in the introduction, the algorithm in Fig. 2 is mathematically equivalent to that suggested in [30]; see Appendix C. It is in order to point out that the material in the appendices relies on \mathbf{I} being volume-preserving and reversible and is otherwise independent of the properties of the velocity Verlet integrator; it is therefore possible to use alternative integrators in Fig. 2, provided that they are both volume-preserving and reversible. For integrators tailored to this task the reader is referred to [6, 23] and their references.

The computation of each step of the chain in Fig. 2 involves two substeps, momentum refreshment \mathbf{r} and dynamics (including an accept/reject mechanism) \mathbf{D} . Fig. 3 describes the computation of the (random) transformation \mathbf{D} . The (random) mapping \mathbf{r} is defined as follows. If $z = (x, y)$, to evaluate $\mathbf{r}(z)$ we draw a realization ζ from the distribution $\mathcal{N}(0, M)$ (see (4)) and set

$$\mathbf{r}(z) = (x, \cos \psi y + \sin \psi \zeta), \quad (6)$$

where ψ is a user specified parameter $0 < \psi \leq \pi/2$. When $\psi = \pi/2$ the effect of \mathbf{r} is to replace the old momentum with a fresh sample of the marginal distribution of y . At the other end of the ψ -interval,

```

00 given  $z_0 = (x_0, y_0) \in \mathbb{R}^d \times \mathbb{R}^d$  and an integer  $N \geq 1$ 
01 for  $n = 0$  to  $N - 1$ 
02      $\bar{z}_{n+1} = \mathbf{r}(z_n)$       % refresh momentum
03      $z_{n+1} = \mathbf{D}(\bar{z}_{n+1})$  % dynamics and accept/reject
04 end for

```

Figure 2: Extra Chance Generalized HMC. It generates a Markov chain $z_0 \rightarrow z_1 \rightarrow \dots \rightarrow z_N$ with ρ in (4) as an invariant density. If y_0 is drawn from the distribution $\mathcal{N}(0, M)$, the marginal chain of positions $x_0 \rightarrow x_1 \rightarrow \dots \rightarrow x_N$ satisfies detailed balance with respect to the target distribution (1).

```

00 given  $z = (x, y) \in \mathbb{R}^d \times \mathbb{R}^d$ 
01  $k = 0$ ,  $\Sigma^{(0)}(z) = 0$ ,  $z^{(0)} = z$ 
02 draw  $u \sim U(0, 1)$       % sample from uniform distribution
03 while  $u > \Sigma^{(k)}$  and  $k < K + 1$ 
04      $z^{(k+1)} = \mathbf{I}(z^{(k)})$  % integration of dynamics
05      $\Sigma^{(k+1)} = \max \left( \Sigma^{(k)}, \min \left( 1, \rho(z^{(k+1)}) / \rho(z^{(0)}) \right) \right)$ 
06      $k \leftarrow k + 1$ 
07 end while
08 if  $u \leq \Sigma^{(k)}$  then
09      $\mathbf{D}(z) = z^{(k)}$       %  $z^{(k)}$  is accepted
10 else
11      $\mathbf{D}(z) = \mathbf{F}(z^{(0)})$  % momentum flip
12 end if

```

Figure 3: Algorithm to compute $\mathbf{D}(z)$. The integer $K \geq 0$ is the number of extra chances, a user-defined parameter.

choosing $\psi \approx 0$ leads to the momentum in $\mathbf{r}(z)$ being a small random perturbation of y as suggested in [15, 17].

The overall algorithm in Fig. 2 has the mass matrix M and the numbers L , Δt , ψ and K as parameters whose values have to be set by the user. In fact, M , L , Δt , and ψ play the same role in XCGHMC as they do in generalized HMC (GHMC); it is not within the scope of the present contribution to discuss the difficult issue of how best to make the corresponding choices. Relevant references are [5, 11] for the choice of M , [4] for Δt , [14] for L , and [15] for ψ . Let us then study the role of K , the number of ‘extra chances.’

As we show presently, with the choice $K = 0$ (no extra chance) the algorithm in Fig. 2 reduces to GHMC [15, 17] and therefore to standard HMC if in addition $\psi = \pi/2$ (complete momentum refreshment). In fact, when $K = 0$, the loop **03–07** in Fig. 3 is executed once per transition $n \rightarrow n + 1$ of the Markov chain. A numerical integration is performed to find the proposed state $z^{(1)} = \mathbf{I}(\bar{z}_{n+1})$ and the next location z_{n+1} of the chain coincides with the proposal $\mathbf{I}(\bar{z}_{n+1})$ (acceptance) with probability

$$\Sigma^{(1)} = \min \left(1, \frac{\rho(\mathbf{I}(\bar{z}_{n+1}))}{\rho(\bar{z}_{n+1})} \right).$$

In the case of rejection, the next location is $z_{n+1} = \mathbf{F}(\bar{z}_{n+1})$. In both cases, acceptance and rejection, the outcome of XCGHMC coincides with that of GHMC. Regardless of the value of ψ , rejection implies that $x_{n+1} = x_n$ and this contributes to an increase of the correlation of the samples; see in this connection the results in [22] and [31]. Accordingly, rejections, while necessary for the algorithm to sample from the right distribution, are not welcome. Furthermore, consider the case where ψ has been chosen small in order that \mathbf{r} almost preserves momentum with the hope that the trajectory z_0, z_1, \dots, z_N approximates the Hamiltonian dynamics (such a choice of ψ is of interest in molecular dynamics simulations). In that case, the fact that a rejection implies reversing the momentum is particularly disturbing, as it clashes with the rationale for the choice of ψ .

Let us now examine the choice $K = 1$ (one extra chance). Consider again the transition $n \rightarrow n + 1$ and assume that for given \bar{z}_{n+1} and u (see Fig. 3), the proposal $\mathbf{I}(\bar{z}_{n+1})$ would have been accepted when using the standard GHMC algorithm. Under this assumption, XCGHMC goes only *once* through the loop **03–07** in Fig. 3 and sets $z_{n+1} = \mathbf{I}(\bar{z}_{n+1})$: both the outcome and the work required by XCGHM coincide with those of GHMC. On the other hand, if for given \bar{z}_{n+1} and u , the proposal $\mathbf{I}(\bar{z}_{n+1})$ would have been rejected in the standard GHMC algorithm, XCGHM offers *one extra chance* of avoiding a momentum flip in the transition $n \rightarrow n + 1$. In fact the loop **03–07** will be executed a second time, which requires an additional integration to find $z^{(2)}$ starting from the point $z^{(1)} = \mathbf{I}(\bar{z}_{n+1})$, and, if $\Sigma^{(2)} \geq u$, the next location will be $z_{n+1} = \mathbf{I}(\mathbf{I}(\bar{z}_{n+1}))$ rather than $\mathbf{F}(\bar{z}_{n+1})$.

The general case $K \geq 1$ should now be clear. In Fig. 3, once a candidate $z^{(k)}$, $k = 1, \dots, K$, has been computed, the integration in line **04** needed to find the next candidate $z^{(k+1)}$ will only be performed if $z^{(k)}$ has not been accepted. A momentum flip will only occur after all candidates $z^{(k)}$, $k = 1, \dots, K + 1$, have failed to be accepted. Thus, with $K \geq 1$, XCGHMC works more than GHMC in *some* transitions $n \rightarrow n + 1$ in an attempt to hopefully avoid the unwelcome outcome $z_{n+1} = \mathbf{F}(\bar{z}_{n+1})$. Since the bulk of the work in the algorithm lies in computing \mathbf{I} , a step of the chain where the candidate $z^{(k)}$, $k = 2, \dots, K + 1$, is accepted costs k times as much as a step in GHMC. Numerical experiments below show that the extra cost per chain step more than pays in terms of sample quality.

The paper [30] uses the words ‘look ahead’ to describe its strategy. We feel this terminology may be misleading as it suggests that in order to accept the proposal $\mathbf{I}(\bar{z}_{n+1})$ it is necessary to look at the points ahead $\mathbf{I}^2(\bar{z}_{n+1}) = \mathbf{I}(\mathbf{I}(\bar{z}_{n+1}))$, $\mathbf{I}^3(\bar{z}_{n+1}) = \mathbf{I}(\mathbf{I}(\mathbf{I}(\bar{z}_{n+1})))$, \dots resulting from taking $2L$, $3L$, \dots leapfrog timesteps from the initial point \bar{z}_{n+1} . The terminology ‘delayed rejection’ used in the statistical literature [12, 20, 32] is also somewhat infelicitous: the goal is to *avoid* rejections, not to make them at a later time.

3. The acceptance probabilities

Let us further study the accept/reject strategy in **D**. Fix z and define (cf. line **05** in Fig. 3) a monotonic sequence

$$0 = \Sigma^{(0)}(z) < \Sigma^{(1)}(z) \leq \dots \leq \Sigma^{(K+1)}(z) \leq 1$$

as follows:

$$\Sigma^{(0)}(z) = 0, \quad \Sigma^{(k)}(z) = \max \left(\Sigma^{(k-1)}(z), \min \left(1, \frac{\rho(\mathbf{I}^k(z))}{\rho(z)} \right) \right), \quad k = 1, \dots, K+1. \quad (7)$$

We emphasize that in this section we are concerned with the *analysis* of **D**: if $k > 1$, the actual algorithm will *not* compute $\Sigma^{(k)}(z)$ (which requires k integration legs with L time-steps each to find $\mathbf{I}^k(z)$) unless the point $\mathbf{I}^{k-1}(z)$ has been rejected (as its predecessors must have). Note that, for $k = 1, \dots, K+1$,

$$\Sigma^{(k)}(z) = \max_{1 \leq j \leq k} \min \left(1, \frac{\rho(\mathbf{I}^j(z))}{\rho(z)} \right) = \min \left(1, \max_{1 \leq j \leq k} \frac{\rho(\mathbf{I}^j(z))}{\rho(z)} \right). \quad (8)$$

From Fig. 3 it is easily concluded that, when computing the random point $\mathbf{D}(z)$, the event ‘ $\mathbf{D}(z)$ is one of $\mathbf{I}(z), \mathbf{I}^2(z), \dots, \mathbf{I}^k(z)$ ’ happens if and only if $u \leq \Sigma^{(k)}(z)$; therefore that event has probability $\Sigma^{(k)}(z)$. As a consequence, the difference

$$p^{(k)}(z) = \Sigma^{(k)}(z) - \Sigma^{(k-1)}(z), \quad k = 1, \dots, K+1 \quad (9)$$

represents the probability that $\mathbf{D}(z) = \mathbf{I}^k(z)$ and

$$p^{(K+2)}(z) = 1 - \Sigma^{(K+1)}(z) \quad (10)$$

is the probability that none among $\mathbf{I}(z), \mathbf{I}^2(z), \dots, \mathbf{I}^{K+1}(z)$ is accepted and therefore $\mathbf{D}(z) = \mathbf{F}(z)$. Pictorially, the values $\Sigma^{(k)}(z)$, $k = 1, \dots, K+1$, partition the interval $[0, 1]$ into $K+2$ subintervals whose lengths $p^{(k)}(z)$, $k = 1, \dots, K+2$, provide the probabilities of the outcomes $\mathbf{D}(z) = \mathbf{I}(z), \dots, \mathbf{D}(z) = \mathbf{I}^{K+1}(z), \mathbf{D}(z) = \mathbf{F}(z)$. Some of the subintervals may degenerate into a single point and then the corresponding outcomes have zero probability.

We note the following conclusions:

1. Assume that $1 < k \leq K+1$ and $\rho(\mathbf{I}^k(z))$ does not exceed the maximum of the preceding $\rho(\mathbf{I}^j(z))$, $j = 1, \dots, k-1$. Then $\Sigma^{(k)}(z) = \Sigma^{(k-1)}(z)$ and the event $\mathbf{D}(z) = \mathbf{I}^k(z)$ has probability $p^{(k)}(z) = 0$: **D** only moves the chain from z to locations $\mathbf{I}^k(z)$ where the density ρ is larger than at all the preceding ‘missed chances’ $\mathbf{I}^j(z)$, $j = 1, \dots, k-1$.
2. Assume that the value of ρ at one location $\mathbf{I}^j(z)$, $1 \leq j \leq k$, exceeds the value $\rho(z)$. Then $\Sigma^{(k)}(z) = 1$ and accordingly one among $\mathbf{I}(z), \mathbf{I}^2(z), \dots, \mathbf{I}^k(z)$ will be accepted. In particular, if $\rho(\mathbf{I}^k(z)) > \rho(z)$ for at least a value of k , $1 \leq k \leq K+1$, the momentum flip $\mathbf{D}(z) = \mathbf{F}(z)$ will not occur.

It is well known that for *symplectic integrators* [13, 18, 27], including the leapfrog scheme, the value of H along a long numerical trajectory of a Hamiltonian system typically oscillates around the value at the initial point of the integration. Here (see (4)) this means that, if K is large, some of the values $\rho(\mathbf{I}^k(z))$ may be expected to be below $\rho(z)$. In view of item 2 above, it then may be hoped that then the flip $\mathbf{D}(z) = \mathbf{F}(z)$ will not occur (see the numerical experiments below).

When comparing the present approach with the literature on delayed rejection [12, 20, 32], it is useful to emphasize that, if z is given, the probabilities $\Sigma^{(k)}(z)$ and $p^{(k)}(z)$ defined above are not conditional. We have not been concerned with questions such as what is the probability of accepting $\mathbf{I}^2(z)$ *conditional* on $\mathbf{I}(z)$ having already been rejected. This is to be compared with the approach in [12, 20, 32], which focuses on the probability of accepting a new attempt conditional on previous attempts at the same step having been rejected. The formulas for such conditional probabilities turn out not to be very neat.

Δt	$K = 0$	$K = 3$				
0.012	93%	93%	6%	1%	0%	100%
0.016	86%	87%	11%	2%	0%	100%
0.020	77%	80%	16%	3%	1%	100%
0.024	65%	71%	22%	6%	1%	100%
		a_0	a_1	a_2	a_3	$\sum a_k$

Table 1: Acceptance rates (rounded to the nearest percentage point) when $\sin \psi = 1$ and $L\Delta t = 0.48$ for different values of Δt and $K = 0$ (standard HMC) and $K = 3$ (three extra chances). The column labelled a_k gives the number corresponding to acceptance taking place after k extra chances.

4. Numerical results

We have implemented the algorithm in Fig. 2 for sampling from the canonical distribution of a molecule of C_9H_{20} . This example, taken from [25], has been used in [7] to compare different Markov chain samplers. It has $3 \times 9 = 27$ degrees of freedom as the hydrogen atoms are lumped to the corresponding carbon atoms. The degrees of freedom may of course be associated with the six rigid body motions and with vibrations in the eight carbon-carbon bond lengths, seven bond angles and six dihedral angles. The potential involves two-, three- and four-atom contributions related to bond lengths, bond angles and dihedral angles respectively and furthermore Lenard-Jones terms for all pairs of atoms separated by three or more covalent bounds. Units and parameter values here are as in [7], with the inverse temperature $\beta = 1$.

The potential energy associated with each dihedral angle has three possible stable equilibria. We are interested in estimating the probability that the absolute value of the dihedral angle ϕ_1 between the first four atoms is below 1.75; this corresponds with ϕ_1 lying in the basin of attraction of the most stable value $\phi_1 = 0$ (for which the first four carbon atoms are coplanar). In other words we assume that the samples x_n are to be used in (2) when A is the indicator of the set $\{|\phi_1| \leq 1.75\}$, i.e. $A = 1$ when $|\phi_1| \leq 1$ and $A = 0$ if $|\phi_1| > 1$. The quality of the samples generated by the different algorithms will be measured by the *effective sample size* (relative to this choice of A); the ESS of a set of $N + 1$ (correlated) Markov chain samples x_n represents the number of *independent* samples that contain the same amount of information; in other words the ESS is the result of dividing $N + 1$ by the autocorrelation time. We estimate ESSs by means of the initial monotone estimator in [10]. Several alternative choices of the observable A were also considered; while ESSs change substantially with A , the corresponding numerical results lead to the same qualitative conclusions as those reported here.

All experiments have a unit mass matrix as in [7]. We studied the cases $K = 0$ (GHMC) and $K = 3$ (three extra chances); the results below show very clearly that it is unnecessary to consider higher values of K . Dozens of combinations of the remaining parameters Δt , L and ψ were tried and we only report here on a representative selection. For each choice of parameters in the algorithm, we generated ten realizations of the chain. The values of the ESS and acceptance probability given here are averages over the ten realizations. Each realization included an initial burn-in phase (500 transitions) and a production phase; the latter comprised 10^6 evaluations of the force $-\nabla V$ so that all realizations use the same amount of computational work, regardless of the value of L . In other words, increasing L results in a lower number N of computed samples. Note also that when $K = 3$ a higher use of extra chances in a realization leads to fewer samples being generated. To complete the description of our experimental setting we mention that, in order to avoid resonances, the value of Δt was slightly randomized at the beginning of each integration leg **I** by subjecting it to a perturbation ranging in a $\pm 5\%$ interval, see e.g. [21].

Table 1 gives values of the acceptance probability when $L\Delta t = 0.48$ and $\sin \psi = 1$ with different values

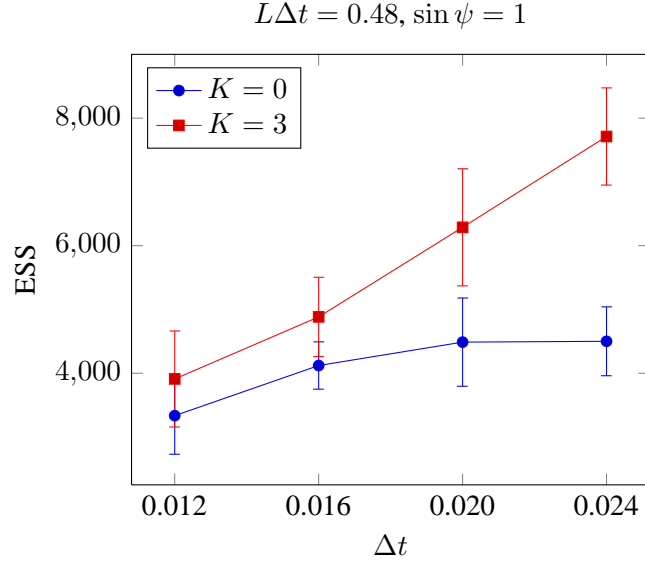


Figure 4: Effective sample size as a function of Δt when the integration interval is $L\Delta t = 0.48$ and $\sin \psi = 1$ (complete momentum refreshment).

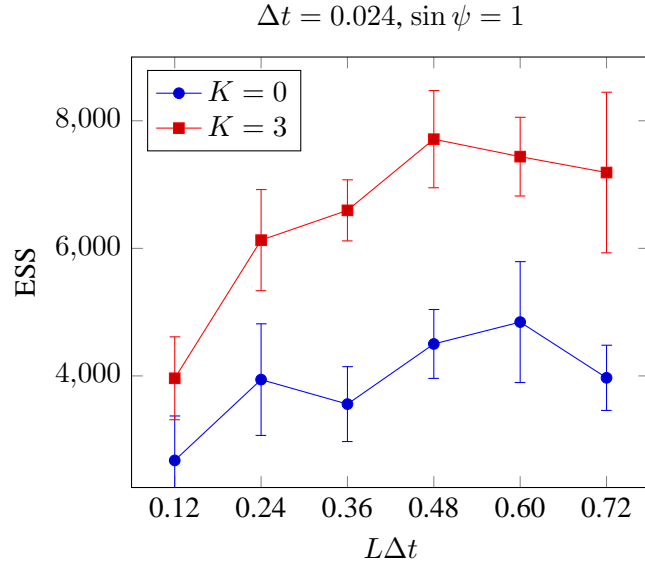


Figure 5: Effective sample size as a function of the time-span $L\Delta t$ of the numerical integration when the time-step is $\Delta t = 0.024$ and $\sin \psi = 1$ (complete momentum refreshment).

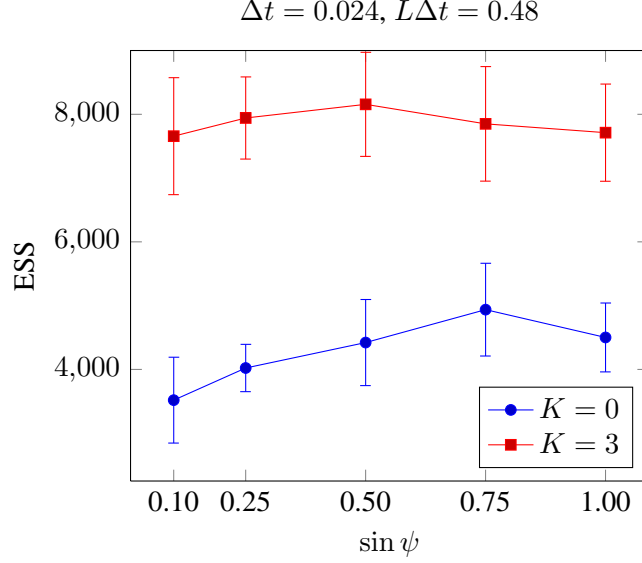


Figure 6: Effective sample size as a function of $\sin \psi$ when the time-step is $\Delta t = 0.024$ and $L\Delta t = 0.48$.

of Δt . The maximum value $\Delta t = 0.024$ considered is fairly close to the upper limit allowed by the size of the stability interval of the Verlet integrator, which experiments suggest is ≈ 0.030 . As expected, for $K = 0$ (standard HMC) the fraction of accepted steps decreases as Δt increases (lower accuracy in the integration, higher energy errors). It is apparent that the use of extra chances almost completely eliminates rejections (in the last row, with two decimal places, $\sum a_k = 99.80\%$). For small Δt the value of a_0 coincides with the acceptance rate for $K = 0$ and in fact, it is clear from our discussion of the algorithm that, when the chains are at *stationarity*, the *expected value* of a_0 exactly matches the expected acceptance rate of HMC. The agreement between the values of a_0 and the acceptance rate for $K = 0$ deteriorates as Δt increases; this happens because for large Δt there are wider differences between the empirical values obtained from the simulations and the corresponding theoretical expected values at stationarity. For large Δt , XGCHMC is less prone to be stuck at the present location than HMC.

Fig. 4 corresponds to the parameter values in Table 1 and gives the ESS of the samples generated. Comparing the values $\Delta t = 0.020$ and $\Delta t = 0.024$ for HMC (i.e. $K = 0$) we see that a lower acceptance rate does not automatically imply a less effective simulation: a larger Δt requires less work per integration leg and provides more samples with a given amount of work. Note that an acceptance rate close to 65% in HMC is sometimes regarded as optimal (see [4] and its references). The figure clearly shows that the additional work per transition required by the extra chances pays: in *all* simulations the ESS improves substantially when moving from $K = 0$ to $K = 3$. In this figure, the best ESS for the extra chance algorithm (7712) is more than 70% higher the best ESS for the its standard counterpart (4501).

In Fig. 5 we study the variation of ESS as a function of the time-span $L\Delta t$ of the numerical integration when $\Delta t = 0.024$ and $\sin \psi = 1$. For $K = 0$, ESS presents a clear dip at $L\Delta t = 0.36$, probably due to a resonance between the integration time-span and the periods involved in the dynamics of ϕ_1 . For all values of $L\Delta t$, the ESS with extra chances improves clearly on that of the standard GHMC. The acceptance rates for the simulations in this figure do not differ much from those given in the last row of Table 1: for instance $L\Delta t = 0.12$ has 67% for $K = 0$ and 74% + 19% + 5% + 1% for $K = 3$.

Finally Fig. 6 shows the dependence of ESS on Horowitz’s angle ψ . As it may have been expected, extra chances are more beneficial for small values of ψ . Again the acceptance rates here are roughly the same as those in the last row of Table 1. We conclude that for fixed Δt the acceptance rates are almost independent of the time-span $L\Delta t$ and ψ .

5. Conclusion

We have studied a technique to avoid rejections/momentum flips in the HMC and GHMC methods. Proof of concept experiments show clearly that the technique is very promising in terms of the number of uncorrelated samples that may be obtained with a given amount of computational work.

The material here may be extended in different directions. It may be combined with more sophisticated integrators [6, 23] (particularly so if the dimensionality of the target is very high and HMC requires small energy errors per degree of freedom) or with the use of shadow Hamiltonians [3, 16]. It may also be easily extended to cover ‘compressible’ variants of HMC such as those presented in [9].

Acknowledgments

We are grateful to M. P. Calvo, M. Girolami, T. Radivojevic, R. D. Skeel and J. Sohl-Dickstein for their inputs and to the ICMAT for the usage of their cluster. This research is supported by Ministerio de Ciencia e Innovación (Spain), under the project MTM2010-18246-C03-01, and by Junta de Castilla y León (Spain) together with the European Social Fund, through a postdoctoral position.

A. Stationarity of the Markov chain

In this appendix we prove that the distribution (4) is invariant for the chain in Fig. 2. Since ρ clearly remains invariant under \mathbf{r} , it is sufficient to show that it also remains invariant under \mathbf{D} . It is well known (see e.g. [9, 19, 26]) that, in turn, such an invariance is implied by the requirement of *modified detailed balance*: for each z and z' ,

$$\rho(z) \rho(z' | z) = \rho(\mathbf{F}(z')) \rho(\mathbf{F}(z) | \mathbf{F}(z')). \quad (\text{A.1})$$

Here $\rho(z' | z)$ denotes the conditional distribution of $z' = \mathbf{D}(z)$ conditional on z and $\rho(\mathbf{F}(z) | \mathbf{F}(z'))$ the conditional distribution of $\mathbf{F}(z) = \mathbf{D}(\mathbf{F}(z'))$ conditional on $\mathbf{F}(z')$ (i.e. on z').

In order to establish (A.1), we shall employ the following identity, which is valid for each z in phase space and $k = 1, \dots, K + 1$ and will be proved at the end of this appendix,

$$\rho(z) p^{(k)}(z) = \rho(\mathbf{F}(\mathbf{I}^k(z))) p^{(k)}(\mathbf{F}(\mathbf{I}^k(z))). \quad (\text{A.2})$$

By definition of \mathbf{D} , the left and right hand-sides of (A.1) have the values

$$\rho(z) \left(\sum_{k=1}^{K+1} p^{(k)}(z) \delta(z' - \mathbf{I}^k(z)) + p^{(K+2)}(z) \delta(z' - \mathbf{F}(z)) \right)$$

and

$$\rho(\mathbf{F}(z')) \left(\sum_{k=1}^{K+1} p^{(k)}(\mathbf{F}(z')) \delta(\mathbf{F}(z) - \mathbf{I}^k(\mathbf{F}(z'))) + p^{(K+2)}(\mathbf{F}(z')) \delta(\mathbf{F}(z) - \mathbf{F}(\mathbf{F}(z'))) \right)$$

$$\begin{array}{ccccccc}
z & \rightarrow & \dots & \rightarrow & \mathbf{I}^j(z) & \rightarrow & \dots & \rightarrow & \mathbf{I}^k(z) \\
\updownarrow \mathbf{F} & & & & \updownarrow \mathbf{F} & & & & \updownarrow \mathbf{F} \\
\mathbf{I}^k(\mathbf{F}(\mathbf{I}^k(z))) & \leftarrow & \dots & \leftarrow & \mathbf{I}^{k-j}(\mathbf{F}(\mathbf{I}^k(z))) & \leftarrow & \dots & \leftarrow & \mathbf{F}(\mathbf{I}^k(z))
\end{array}$$

Figure A.7: Top row: k successive applications of \mathbf{I} starting from the point z . Bottom row: k successive applications of \mathbf{I} starting from $\mathbf{F}(\mathbf{I}^k(z))$. The reversibility of \mathbf{I} implies that each point in the bottom row may be obtained from the corresponding point in the top row by flipping the momentum. Therefore each point in the bottom row possesses the same density ρ as the corresponding point above.

respectively. Since the last terms in these expressions obviously coincide, the proof will be over if we show that, for $k = 1, \dots, K + 1$,

$$\rho(z) p^{(k)}(z) \delta(z' - \mathbf{I}^k(z)) = \rho(\mathbf{F}(z')) p^{(k)}(\mathbf{F}(z')) \delta(\mathbf{F}(z) - \mathbf{I}^k(\mathbf{F}(z'))).$$

This is a consequence of (A.2), as the reversibility of \mathbf{I} implies

$$\delta(\mathbf{F}(z) - \mathbf{I}^k(\mathbf{F}(z'))) = \delta(z - \mathbf{F}(\mathbf{I}^k(\mathbf{F}(z')))) = \delta(z' - \mathbf{I}^k(z))$$

because both \mathbf{F} and \mathbf{I}^k are volume preserving. The proof of (A.1) is then ready.

Let us finally prove the identity (A.2). We restrict the attention to the case $k > 1$; the proof for the case $k = 1$ is similar but simpler. After using the expression for $p^{(k)}$ in (10), we have to show that

$$\begin{aligned}
& \min \left(\rho(z), \max_{1 \leq j \leq k} \rho(\mathbf{I}^j(z)) \right) - \min \left(\rho(z), \max_{1 \leq j \leq k-1} \rho(\mathbf{I}^j(z)) \right) \\
&= \min \left(\rho(\mathbf{F}(\mathbf{I}^k(z))), \max_{1 \leq j \leq k} \rho(\mathbf{I}^j(\mathbf{F}(\mathbf{I}^k(z)))) \right) - \min \left(\rho(\mathbf{F}(\mathbf{I}^k(z))), \max_{1 \leq j \leq k-1} \rho(\mathbf{I}^j(\mathbf{F}(\mathbf{I}^k(z)))) \right).
\end{aligned}$$

The reversibility of \mathbf{I} implies (see Fig. A.7) that the values of ρ that feature on the right hand-side are actually the same as those appearing in the left hand side. In fact, if \hat{j} , $1 \leq \hat{j} \leq k - 1$, is such that

$$\rho(\mathbf{I}^{\hat{j}}(z)) = \max_{1 \leq j \leq k-1} \rho(\mathbf{I}^j(z)),$$

the equality to be established reads

$$\begin{aligned}
& \min \left(\rho(z), \max \left(\rho(\mathbf{I}^{\hat{j}}(z)), \rho(\mathbf{I}^k(z)) \right) \right) - \min \left(\rho(z), \rho(\mathbf{I}^{\hat{j}}(z)) \right) \\
&= \min \left(\rho(\mathbf{I}^k(z)), \max \left(\rho(z), \rho(\mathbf{I}^{\hat{j}}(z)) \right) \right) - \min \left(\rho(\mathbf{I}^k(z)), \rho(\mathbf{I}^{\hat{j}}(z)) \right).
\end{aligned}$$

This is checked by successively considering the six possible orderings $\rho(z) \leq \rho(\mathbf{I}^{\hat{j}}(z)) \leq \rho(\mathbf{I}^k(z))$, $\rho(z) \leq \rho(\mathbf{I}^k(z)) \leq \rho(\mathbf{I}^{\hat{j}}(z))$, etc.

B. Detailed balance

While, as proved above, a single step of \mathbf{D} satisfies modified detailed balance, the same is not true for the whole chain in Fig. 2. The paper [9] presents a relevant counterexample and also shows that modified detailed balance is achieved by imposing a palindromic structure to the transitions $n \rightarrow n + 1$ as in Fig. B.8.

```

00  given  $Z_0 = (X_0, Y_0) \in \mathbb{R}^d \times \mathbb{R}^d$  and an integer  $N \geq 1$ 
01  for  $n = 0$  to  $N - 1$ 
02       $\bar{Z}_{n+1} = \mathbf{R}(Z_n)$  % refresh momentum
03       $\hat{Z}_{n+1} = \mathbf{D}(\bar{Z}_{n+1})$  % dynamics and accept/reject
04       $Z_{n+1} = \mathbf{R}(\hat{Z}_{n+1})$  % refresh momentum
05  end for

```

Figure B.8: A ‘tought experiment’ chain that satisfies modified detailed balance with respect to ρ in (4).

The lines **02** and **04** refresh the momentum as in (6); the use of a capital **R** indicates that Fig. B.8 may employ an angle Ψ , $0 < \Psi \leq \pi/2$ different from that ψ in Fig. 2:

$$\mathbf{R}(Z) = (X, \cos \Psi Y + \sin \Psi \zeta).$$

In fact in what follows we assume that Ψ is determined as a function of ψ through the relation $\cos^2 \Psi = \cos \psi$.

We show next that the XCGHMC chain (Fig. 2) implemented in practice and the ‘tought experiment chain’ in Fig. B.8 give rise to same marginal chain for the positions x_n or X_n and that, therefore, the chain of positions in XCGHMC is reversible with respect to the target (1).

Let us consider an arbitrary realization of the auxiliary chain in Fig. B.8. This will correspond to a starting location (X_0, Y_0) (with Y_0 drawn from the distribution $\mathcal{N}(0, M)$) and to realizations $\bar{\zeta}_{n+1}$, u_{n+1} , $\hat{\zeta}_{n+1}$ of the random variables used in steps **02**, **03** and **04** respectively. We define a realization of the XCGHMC chain by setting $x_0 = X_0$,

$$\begin{aligned}
y_0 &= \cos(\psi - \Psi) Y_0 - \sin(\psi - \Psi) \bar{\zeta}_1, \\
\zeta_1 &= \sin(\psi - \Psi) Y_0 + \cos(\psi - \Psi) \bar{\zeta}_1, \\
\zeta_{n+1} &= (1/\sin \psi)(\cos \Psi \sin \Psi \hat{\zeta}_n + \sin \Psi \bar{\zeta}_{n+1}), \quad n = 1, \dots, N-1,
\end{aligned}$$

and using ζ_{n+1} and u_{n+1} , $n = 0, \dots, N-1$, when computing **r** and **D** in steps **02** and **03** of Fig. 2 respectively. Note that this is legitimate because, as it is easily checked, the random variables ζ_n possess the right $\mathcal{N}(0, M)$ distribution. The definitions of y_0 and ζ_1 result in $\bar{z}_1 = \bar{Z}_1$, which in turn implies that $z_1 = \hat{Z}_1$. Then the choice of ζ_2 ensures that the momentum refreshment to get $\bar{z}_2 = \mathbf{r}(z_1)$ produces the same output as the two momentum refreshments to compute $\bar{Z}_2 = \mathbf{R}(Z_1) = \mathbf{R}(\mathbf{R}(\hat{Z}_1))$. The iteration of this argument shows that, for $n = 1, \dots, N$, $z_n = \hat{Z}_n$ so that $x_n = X_n$. We conclude that we may regard the samples x_n as originating from the palindromic chain in Fig. B.8.

C. Equivalence of XCGHMC and the algorithm by Sohl-Dickstein *et al.*

In [30], the authors introduced an algorithm called Look Ahead Hybrid Monte Carlo (LAHMC) which is in fact equivalent to XCGHMC. There, the transition probabilities to $z^{(k)}$ are defined as follows

$$\pi^{(k)}(z) = \min \left(1 - \sum_{1 \leq j \leq k-1} \pi^{(j)}(z), \frac{\rho(\mathbf{F} \mathbf{I}^k(z))}{\rho(z)} \left(1 - \sum_{1 \leq j \leq k-1} \pi^{(j)}(\mathbf{F} \mathbf{I}^k(z)) \right) \right), \quad k = 1, \dots, K+1. \tag{C.1}$$

In order to show that XCGHMC and LAHMC are equivalent, it suffices to show that the accumulated probabilities

$$S^{(0)}(z) = 0, \quad S^{(k)}(z) = \sum_{1 \leq j \leq k} \pi^{(j)}(z), \quad k = 1, \dots, K+1. \quad (\text{C.2})$$

agree with $\Sigma^{(k)}$ in (8) for all k at any point z . Simple manipulations prove that this is true for $k = 1$; the general case is proved by induction. Assume the assertion true up to some $1 \leq k-1 \leq K$ and let us show that it holds for k . To lighten the writing, we set

$$M^{(k)}(z) = \max_{1 \leq j \leq k} \frac{\rho(\mathbf{F} \mathbf{I}^j(z))}{\rho(z)} = \max_{1 \leq j \leq k} \frac{\rho(\mathbf{I}^j(z))}{\rho(z)}.$$

By definition,

$$\begin{aligned} S^{(k)}(z) &= S^{(k-1)}(z) + \min \left(1 - S^{(k-1)}(z), \frac{\rho(\mathbf{F} \mathbf{I}^k(z))}{\rho(z)} \left(1 - S^{(k-1)}(\mathbf{F} \mathbf{I}^k(z)) \right) \right) \\ &= \min \left(1, \frac{\rho(\mathbf{F} \mathbf{I}^k(z))}{\rho(z)} \left(1 - S^{(k-1)}(\mathbf{F} \mathbf{I}^k(z)) \right) + S^{(k-1)}(z) \right) \\ &= \min \left(1, \frac{\rho(\mathbf{F} \mathbf{I}^k(z))}{\rho(z)} \left(1 - \min(1, M^{(k-1)}(\mathbf{F} \mathbf{I}^k(z))) \right) + S^{(k-1)}(z) \right) \\ &= \min \left(1, \max \left(0, \frac{\rho(\mathbf{F} \mathbf{I}^k(z))}{\rho(z)} - M^{(k-1)}(z) \right) + S^{(k-1)}(z) \right), \end{aligned}$$

where we have used the hypothesis of induction on $S^{(k-1)}(\mathbf{F} \mathbf{I}^k(z))$ at the third equality and the fact that

$$\rho(z) M^{(k-1)}(z) = \rho(\mathbf{F} \mathbf{I}^k(z)) M^{(k-1)}(\mathbf{F} \mathbf{I}^k(z))$$

at the last equality (recall that ρ is invariant under \mathbf{F} and that $\mathbf{I}^j \mathbf{F} \mathbf{I}^k = \mathbf{I}^{k-j}$).

Now, two possibilities arise, either $S^{(k-1)}(z) < 1$ or $S^{(k-1)}(z) = 1$. If $S^{(k-1)}(z) < 1$, in which case by the hypothesis of induction $S^{(k-1)}(z) = M^{(k-1)}(z)$, then

$$S^{(k)}(z) = \min \left(1, \max \left(M^{(k-1)}(z), \frac{\rho(\mathbf{F} \mathbf{I}^k(z))}{\rho(z)} \right) \right).$$

If on the contrary $S^{(k-1)}(z) = 1$, the hypothesis of induction implies that $M^{(k-1)}(z) \geq 1$, hence

$$S^{(k)}(z) = 1 = \min \left(1, \max \left(M^{(k-1)}(z), \frac{\rho(\mathbf{F} \mathbf{I}^k(z))}{\rho(z)} \right) \right).$$

Therefore, in both cases

$$S^{(k)}(z) = \min \left(1, \max \left(M^{(k-1)}(z), \frac{\rho(\mathbf{F} \mathbf{I}^k(z))}{\rho(z)} \right) \right) = \min \left(1, M^{(k)}(z) \right),$$

and the proof is complete.

In our opinion working with accumulated probabilities as in XCGHMC rather than with the probabilities $\pi^{(k)}$ not only leads to better theoretical insights as in Section 3, but results in simpler code.

References

- [1] Akhmatskaya, E., Bou-Rabee, N., Reich, S., 2009. A comparison of generalized hybrid Monte Carlo methods with and without momentum flip. *J. Comput. Phys.* 228 (6), 2256–2265.
URL <http://dx.doi.org/10.1016/j.jcp.2008.12.014>
- [2] Akhmatskaya, E., Bou-Rabee, N., Reich, S., 2009. Erratum to “A comparison of generalized hybrid Monte Carlo methods with and without momentum flip” [*J. Comput. Phys.* 228 (2009) 2256–2265] [mr2500680]. *J. Comput. Phys.* 228 (19), 7492–7496.
URL <http://dx.doi.org/10.1016/j.jcp.2009.06.039>
- [3] Akhmatskaya, E., Reich, S., 2008. GSHMC: an efficient method for molecular simulation. *J. Comput. Phys.* 227 (10), 4934–4954.
URL <http://dx.doi.org/10.1016/j.jcp.2008.01.023>
- [4] Beskos, A., Pillai, N., Roberts, G., Sanz-Serna, J.-M., Stuart, A., 2013. Optimal tuning of the hybrid Monte Carlo algorithm. *Bernoulli* 19 (5A), 1501–1534.
URL <http://dx.doi.org/10.3150/12-BEJ414>
- [5] Beskos, A., Pinski, F. J., Sanz-Serna, J. M., Stuart, A. M., 2011. Hybrid Monte Carlo on Hilbert spaces. *Stochastic Process. Appl.* 121 (10), 2201–2230.
URL <http://dx.doi.org/10.1016/j.spa.2011.06.003>
- [6] Blanes, S., Casas, F., Sanz-Serna, J. M., 2014. Numerical Integrators for the Hybrid Monte Carlo Method. *SIAM J. Sci. Comput.* 36 (4), A1556–A1580.
URL <http://dx.doi.org/10.1137/130932740>
- [7] Cancès, E., Legoll, F., Stoltz, G., 2007. Theoretical and numerical comparison of some sampling methods for molecular dynamics. *M2AN Math. Model. Numer. Anal.* 41 (2), 351–389.
URL <http://dx.doi.org/10.1051/m2an:2007014>
- [8] Duane, S., Kennedy, A. D., Pendleton, B. J., Roweth, D., 1987. Hybrid Monte Carlo. *Phys. Lett. B* 195 (2), 216–222.
URL [http://dx.doi.org/10.1016/0370-2693\(87\)91197-X](http://dx.doi.org/10.1016/0370-2693(87)91197-X)
- [9] Fang, Y., Sanz-Serna, J. M., Skeel, R. D., 2014. Compressible generalized hybrid Monte Carlo. *J. Chem. Phys.* 140 (17), 174109.
URL <http://dx.doi.org/10.1063/1.4874000>
- [10] Geyer, C. J., 1992. Practical Markov Chain Monte Carlo. *Statist. Sci.* 7 (4), 473–483.
URL <http://dx.doi.org/10.1214/ss/1177011137>
- [11] Girolami, M., Calderhead, B., 2011. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 73 (2), 123–214, with discussion and a reply by the authors.
URL <http://dx.doi.org/10.1111/j.1467-9868.2010.00765.x>
- [12] Green, P. J., Mira, A., 2001. Delayed rejection in reversible jump Metropolis-Hastings. *Biometrika* 88 (4), 1035–1053.
URL <http://dx.doi.org/10.1093/biomet/88.4.1035>
- [13] Hairer, E., Lubich, C., Wanner, G., 2006. Geometric numerical integration, 2nd Edition. Vol. 31 of Springer Series in Computational Mathematics. Springer-Verlag, Berlin, structure-preserving algorithms for ordinary differential equations.
URL <http://dx.doi.org/10.1007/3-540-30666-8>
- [14] Hoffman, M. D., Gelman, A., 2014. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. R.* 15, 1593–1623.
URL <http://jmlr.org/papers/v15/hoffman14a.html>
- [15] Horowitz, A. M., 1991. A generalized guided Monte Carlo algorithm. *Phys. Lett. B* 268 (2), 247–252.
URL [http://dx.doi.org/10.1016/0370-2693\(91\)90812-5](http://dx.doi.org/10.1016/0370-2693(91)90812-5)
- [16] Izaguirre, J. A., Hampton, S. S., 2004. Shadow hybrid Monte Carlo: an efficient propagator in phase space of macromolecules. *J. Comput. Phys.* 200 (2), 581–604.
URL <http://dx.doi.org/10.1016/j.jcp.2004.04.016>
- [17] Kennedy, A. D., Pendleton, B., 2001. Cost of the generalised hybrid Monte Carlo algorithm for free field theory. *Nuclear Phys. B* 607 (3), 456–510.
URL [http://dx.doi.org/10.1016/S0550-3213\(01\)00129-8](http://dx.doi.org/10.1016/S0550-3213(01)00129-8)
- [18] Leimkuhler, B., Reich, S., 2004. Simulating Hamiltonian dynamics. Vol. 14 of Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, Cambridge.
URL <http://dx.doi.org/10.1017/CBO9780511614118>
- [19] Lelièvre, T., Rousset, M., Stoltz, G., 2010. Free energy computations. Imperial College Press, London, a mathematical perspective.
URL <http://dx.doi.org/10.1142/9781848162488>

- [20] Mira, A., 2001. On Metropolis-Hastings algorithms with delayed rejection. *Metron* LIX (3-4), 231–241.
URL <http://EconPapers.repec.org/RePEc:mtn:ancoec:2001:3:16>
- [21] Neal, R. M., 2011. MCMC using Hamiltonian dynamics. In: *Handbook of Markov chain Monte Carlo*. Chapman & Hall/CRC Handb. Mod. Stat. Methods. CRC Press, Boca Raton, FL, pp. 113–162.
URL <http://www.mcmchandbook.net/HandbookChapter5.pdf>
- [22] Peskun, P. H., 1973. Optimum Monte-Carlo sampling using Markov chains. *Biometrika* 60, 607–612.
URL <http://www.jstor.org/stable/2335011>
- [23] Predescu, C., Lippert, R. A., Eastwood, M. P., Ierardi, D., Xu, H., Jensen, M. O., Bowers, K. J., Gullingsrud, J., Rendleman, C. A., Dror, R. O., Shaw, D. E., 2012. Computationally efficient molecular dynamics integrators with improved sampling accuracy. *Mol. Phys.* 110 (9-10), 967–983.
URL <http://dx.doi.org/10.1080/00268976.2012.681311>
- [24] Robert, C. P., Casella, G., 2004. *Monte Carlo statistical methods*, 2nd Edition. Springer Texts in Statistics. Springer-Verlag, New York.
URL <http://dx.doi.org/10.1007/978-1-4757-4145-2>
- [25] Ryckaert, J.-P., Bellemans, A., 1978. Molecular dynamics of liquid alkanes. *Faraday Discuss. Chem. Soc.* 66 (0), 95–106.
URL <http://dx.doi.org/10.1039/DC9786600095>
- [26] Sanz-Serna, J. M., 2014. Markov Chain Monte Carlo and numerical differential equations. In: *Current Challenges in Stability Issues for Numerical Differential Equations*. Vol. 2082 of *Lecture Notes in Mathematics*. Springer, pp. 39–88.
URL <http://www.springer.com/978-3-319-01299-5>
- [27] Sanz-Serna, J. M., Calvo, M. P., 1994. *Numerical Hamiltonian problems*. Vol. 7 of *Applied Mathematics and Mathematical Computation*. Chapman & Hall, London.
URL <http://dx.doi.org/10.1007/978-1-4899-3093-4>
- [28] Schlick, T., 2010. *Molecular modeling and simulation*, 2nd Edition. Vol. 21 of *Interdisciplinary Applied Mathematics*. Springer, New York, an interdisciplinary guide.
URL <http://dx.doi.org/10.1007/978-1-4419-6351-2>
- [29] Sohl-Dickstein, J., 2012. Hamiltonian Monte Carlo with Reduced Momentum Flips, arXiv:1205.1939.
URL <http://arxiv.org/abs/1205.1939>
- [30] Sohl-Dickstein, J., Mudigonda, M., DeWeese, M. R., 2014. Hamiltonian Monte Carlo Without Detailed Balance. In: *ICML*. Vol. 32 of *JMLR Proceedings*. JMLR.org, pp. 719–726.
URL <http://jmlr.org/proceedings/papers/v32/sohl-dickstein14.html>
- [31] Tierney, L., 1998. A note on Metropolis-Hastings kernels for general state spaces. *Ann. Appl. Probab.* 8 (1), 1–9.
URL <http://dx.doi.org/10.1214/aoap/1027961031>
- [32] Tierney, L., Mira, A., 1999. Some adaptive Monte Carlo methods for Bayesian inference. *Statist. Med.* 18 (17-18), 2507–2515.
URL [http://dx.doi.org/10.1002/\(SICI\)1097-0258\(19990915/30\)18:17/18<2507::AID-SIM272>3.0.CO;2-J](http://dx.doi.org/10.1002/(SICI)1097-0258(19990915/30)18:17/18<2507::AID-SIM272>3.0.CO;2-J)
- [33] Wagoner, J. A., Pande, V. S., 2012. Reducing the effect of Metropolization on mixing times in molecular dynamics simulations. *J. Chem. Phys.* 137 (21), 214105.
URL <http://dx.doi.org/10.1063/1.4769301>